HAL

archives-ouvertes.fr

Statistical inverse problems for non-Gaussian vector valued random fields with a set of experimental realizations

G. Perrin, Christian Soize, Denis Duhamel, C. Fünfschilling

▶ **To cite this version:**

HAL Id: hal-00806417

https://hal-upec-upem.archives-ouvertes.fr/hal-00806417

Submitted on 30 Mar 2013

Statistical inverse problems for non-Gaussian vector valued random fields with a set of experimental realizations

G. Perrin & C. Soize
Modélisation et Simulation Multi-Échelle (MSME UMR 8208 CNRS)
Université Paris-Est, Marne-la-Vallée, France.

D. Duhamel
Navier (UMR 8205 ENPC-IFSTTAR-CNRS)
Université Paris-Est, Marne-la-Vallée, France.

C. Funfschilling
Innovation and Research Department
SNCF, Paris, France.

ABSTRACT: The railway track irregularities, which can be seen as a four dimensions vector-valued random field, are the main source of excitation of the train, which is a very nonlinear system with many degrees-of-freedom. Due to the specific interaction between the track and the train, this random field is neither Gaussian nor stationary. In order to characterize the train dynamics, a complete parametrization of the physical and statistical properties of the track irregularities would thus be of great concern. Based on a set of track measurements, which can be seen as a set of independent realizations, this work proposes an innovative approach to optimally identify in inverse the characteristics of the track irregularities random field.

At first, using a revisited Karhunen-Loève expansion, the considered random field is approximated by its truncated projection on a particularly well adapted orthogonal basis. Then, the distribution of the random vector that gathers the projection coefficients of the random field on this spatial basis is characterized using a polynomial chaos expansion. The dimension of this random vector being very high (around five hundred), advanced identification techniques are introduced to allow performing relevant convergence analysis and identification.

Based on the stochastic modeling of the non- Gaussian non-stationary vector-valued track geometry random field, realistic track geometries, which are representative of the experimental measurements and representative of the whole railway network, can be generated. These tracks can then be introduced as an input of any railway software to characterize the stochastic behavior of any normalized train.

1 INTRODUCTION

The expected benefits of simulation in the railway field are multiple: robust and optimized conception, shorter and cheaper certification procedure, better knowledge of the critical situations of the track/vehicle system, optimization of the maintenance.

At its building, the theoretical new railway line is supposed to be made of perfect straight lines and curves. This track geometry is however gradually damaged and regularly subjected to maintenance operations. Therefore, during their lifecycles, trains are confronted to very different running conditions.

If simulation is introduced in certification and con-ception processes, it has to be very representative of the physical behaviour of the system: the model of the train has to be fully validated and the simulations have to be raised on a realistic and representative set of track geometries. This requires the development of a complete stochastic modeling of the track irregularities, for which the spatial and statistical dependencies between the different track irregularities have to be accurately taken into account.

In reply to these expectations, the track geometry of the French railway network has been continuously measured since 2007. Due to the specific interaction between the train and track, this track irregularity random field is neiher Gaussian nor stationary, which makes the identification of the distribution of the track

irregularity tough. Based on the experimental measurements, which can be seen as a finite set of independent realizations, the work proposes an innovative method to completely parametrize the track irregularity random field.

First, this method is based on a revisited Karhunen-Loève (KL) decomposition, which is presented in Section 2. This original decomposion makes a point of maximizing the representativeness of the projection basis with respect to the available information. Section 3 deals then with the second step of the modeling, which is an inverse polynomial chaos expansion identification. At last, Section 4 shows in what extent such an approach allows the generation of track geometries that are similar to measured ones from a frequency and statistical point of view.

2 OPTIMAL REDUCED BASIS

Let $\mathbb{H} = L^2(\Omega, \mathbb{R}^4)$ be the space of square integrable functions on $\Omega = [0, S]$, with values in $\mathbb{R}^4$, equipped with the inner product $(\cdot, \cdot)$, such that for all $\boldsymbol{z}^1$ and $\boldsymbol{z}^2$ in $\mathbb{H}$:

$$\left(\boldsymbol{z}^1, \boldsymbol{z}^2\right) = \int_\Omega \left(\boldsymbol{z}^1(s)\right)^T \boldsymbol{z}^2(s) ds. \tag{1}$$

The track irregularity random field is denoted by

$$\{\boldsymbol{X}(s) = (X_1(s), X_2(s), X_3(s), X_4(s)),\ s \in \Omega\}, \tag{2}$$

where s is the curvilinear abscissia of the track, $\Omega = [0, S]$ is a bounded interval and X_1, X_2, X_3 and X_4 refer to the four kinds of track irregularity that are strongly dependent.

It is supposed that $\boldsymbol{X}$ is a second order centered random field, such that:

$$E[\boldsymbol{X}(s)] = 0,\ \forall s \in \Omega,\ \ \|\boldsymbol{X}\|_2^2 \stackrel{\text{def}}{=} E[(\boldsymbol{Z}, \boldsymbol{Z})] < +\infty, \tag{3}$$

where $E[\cdot]$ is the mathematical expectation.

It is assumed in this work that ν independent realizations of $\boldsymbol{X}$, $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^\nu\}$, have been measured. This defines the maximum available information about $\boldsymbol{X}$ for the modeling.

For any $M \geq 1$, based on this available data, the idea of this section is to identify the M-dimension orthonormal projection family $\mathcal{F}_{opt}^{(M)} = \{\boldsymbol{f}^k,\ 1 \leq k \leq M\} \in \mathbb{H}^M$, such that:

$$\boldsymbol{X} \approx \widehat{\boldsymbol{X}}^{(M)}(\mathcal{F}_{opt}^{(M)}) \stackrel{\text{def}}{=} \sum_{i=k}^M \boldsymbol{f}^k C_k,\ \left(\boldsymbol{f}^k, \boldsymbol{f}^\ell\right) = \delta_{k\ell}, \tag{4}$$

with $\delta_{k\ell}$ is the kronecker symbol equal to one if $k = \ell$ and zero otherwise, which minimizes, among all the families $\mathcal{F}^{(M)}$ in $\mathbb{H}^M$, the maximal normalized amplitude of the truncation residual of each track irregularity:

$$\mathcal{F}_{opt}^{(M)} = \arg \min_{\mathcal{F}^{(M)} \in \mathbb{H}^M} \left\{ \max_{1 \leq i \leq 4} \frac{\left\| X_i - \widehat{X}_i^{(M)}(\mathcal{F}_{opt}^{(M)}) \right\|_2^2}{\|X_i\|_2^2} \right\}. \tag{5}$$

To this end, an original expansion is proposed in this section, that is based on a double adaptation of the classical KL expansion.

2.1 *Optimality of the KL decomposition*

For any M-dimensional set of functions of $\mathbb{H}$, $\mathcal{F}^{(M)}$, we define $\widehat{\boldsymbol{X}}(\mathcal{F}^{(M)})$ to denote the projection of $\boldsymbol{X}$ on $\mathcal{F}^{(M)}$. The matrix-valued covariance function, $[R_{XX}]$, of centered random field $\boldsymbol{X}$ is moreover written:

$$[R_{XX}(s, s')] = E[\boldsymbol{X}(s) \otimes \boldsymbol{X}(s')],\ (s, s') \in \Omega^2. \tag{6}$$

The KL basis, $\mathcal{K} = \{\boldsymbol{b}^k,\ 1 \leq k\}$, associated with $\boldsymbol{X}$, can be constructed as a countable Hilbertian basis of $\mathbb{H}$, which is constituted of the eigenfunctions of covariance matrix-valued function $[R_{\boldsymbol{XX}}]$, such that for s in Ω and $k, \ell \geq 1$:

$$\int_\Omega [R_{\boldsymbol{XX}}(s, s')] \boldsymbol{b}^k(s') ds' = \lambda_k \boldsymbol{b}^k(s), \tag{7}$$

$$\left(\boldsymbol{b}^k, \boldsymbol{b}^\ell\right) = \delta_{k\ell}, \quad \lambda_1 \geq \lambda_2 \geq \cdots \to 0. \tag{8}$$

Equation (7) is usually called Fredholm problem (see Le Maître and Knio 2010, Hansen 1992, Weese 1992 for further details about its solving). Due to the orthogonal projection theorem in Hilbert space, for $M \geq 1$, projection family $\mathcal{K}^{(M)} = \{\boldsymbol{b}^k,\ 1 \leq k \leq M\}$ is optimal in the sense that, for all family $\mathcal{F}^{(M)}$ in $\mathbb{H}^M$:

$$\left\| \boldsymbol{X} - \widehat{\boldsymbol{X}}(\mathcal{K}^{(M)}) \right\|_2^2 \leq \left\| \boldsymbol{X} - \widehat{\boldsymbol{X}}(\mathcal{F}^{(M)}) \right\|_2^2. \tag{9}$$

2.2 *Scaled Karhunen-Loève expansion*

Noticing that:

$$\left\| \boldsymbol{X} - \widehat{\boldsymbol{X}}(\mathcal{K}^{(M)}) \right\|_2^2 = \sum_{i=1}^{4} \|X_i\|_2^2 \frac{\left\| X_i - \widehat{X}_i(\mathcal{K}^{(M)}) \right\|_2^2}{\|X_i\|_2^2},$$

(10)

it can be seen that the KL expansion minimizes in priority the normalized amplitudes of the residues corresponding to the components of $\boldsymbol{X}$ that have the highest signal energies, $\|X_i\|_2^2$.

In Perrin et al. 2013, it has been shown that this can lead to cases where the components of $\boldsymbol{X}$ whose signal energies are low can be very badly characterized, although error $\left\| \boldsymbol{X} - \widehat{\boldsymbol{X}}(\mathcal{K}^{(M)}) \right\|_2^2$ is low. In order to avoid this phenomenon, a scaled KL decomposition has thus been proposed. This innovative decomposition is based on the KL expansion of an other random field, $\boldsymbol{Y}(\boldsymbol{O}) = [O]\boldsymbol{X}$, where $[O]$ is the following diagonal matrix:

$$[O] = \begin{bmatrix} O_1 & 0 & 0 & 0 \\ 0 & O_2 & 0 & 0 \\ 0 & 0 & O_3 & 0 \\ 0 & 0 & 0 & O_4 \end{bmatrix},$$

(11)

and where $\boldsymbol{O} = (O_1, O_2, O_3, O_4)$ is a deterministic vector which belongs to

$$\mathcal{O} = \left\{ \boldsymbol{x} \in]0,1[^4, \ \|\boldsymbol{x}\|^2 = 1 \right\}.$$

(12)

Therefore, Perrin et al. 2013 showed that the optimization problem defined by Eq. (5) can be approximated by an other optimization problem over $\boldsymbol{O}$ in $\mathcal{O}$:

$$\mathcal{F}_{opt}^{(M)} \approx \mathcal{F}_{+\infty}^{(M)},$$

(13)

$$\mathcal{F}_{+\infty}^{(M)} = \arg \min_{\boldsymbol{O} \in \mathcal{O}} \left\{ \max_{1 \leq i \leq 4} \frac{\left\| Y_i(\boldsymbol{O}) - \widehat{Y}_i(\boldsymbol{O}, \mathcal{F}^{(M)}) \right\|_2^2}{\|Y_i(\boldsymbol{O})\|_2^2} \right\}.$$

(14)

A very efficient iterative algorithm is moreover given in this work to solve Eq. (14).

Finally, when interested in studying complex systems that are excited by vector-valued random fields, which is the case of the interaction between the trains and the track irregularities, the scaled KL expansion opens very interesting opportunities to identify projection basis that do not defavorize the characterization of a particular component of $\boldsymbol{X}$, which could have the lowest signal energy but play the most important role on the mechanical response of the system.

2.3 *Optimal reduced basis from a finite set of realizations*

In Section 2.2, for $M \geq 1$, a method to identify optimal projection family $\mathcal{F}_{opt}^{(M)}$ has been presented, which is the original goal of this section. This method is based on the knowledge of the covariance matrix-valued function $[R_{\boldsymbol{X}\boldsymbol{X}}]$. As the maximal available information about $\boldsymbol{X}$ is characterized by a set of independent realizations, $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^\nu\}$, this covariance function can however not be exactly identified, and its *a priori* best evaluation is given by the following empirical estimator, for (s, s') in $\Omega \times \Omega$:

$$[R_{XX}(s,s')] \approx [\widehat{R}_{XX}(s,s')] \overset{\text{def}}{=} \frac{1}{\nu} \sum_{n=1}^{\nu} x^n(s) x^n(s').$$

(15)

Nevertheless, there is no reason for the projection basis associated with $[\widehat{R}_{XX}]$ to be still optimal with respect to the minimization of error

$$\varepsilon^2(\mathcal{F}^{(M)}) \overset{\text{def}}{=} \left\| \boldsymbol{X} - \widehat{\boldsymbol{X}}(\mathcal{F}^{(M)}) \right\|_2^2, \ \ \mathcal{F}^{(M)} \in \mathbb{H}^M. \ \ (16)$$

The KL expansion being optimal for $\boldsymbol{X}$ with respect to the minimization of error ε^2, it can however be deduced that reciprocally, for all $M \geq 1$:

$$[R_{XX}] = \arg \min_{[A] \in \mathcal{S}(\mathbb{R})} \left\{ \varepsilon^2(\mathcal{B}_{[A]}^{(M)}) \right\},$$

(17)

where:

$$\mathcal{S}(\mathbb{R}) = \big\{ [A] \in L^2(\Omega \times \Omega, \mathbb{R}^4 \times \mathbb{R}^4), \ |$$

$$[A(s,s')] = [A(s',s)], \ (s,s') \in \Omega \times \Omega \big\},$$

(18)

and where for all $[A]$ in $\mathcal{S}(\mathbb{R})$, $\mathcal{B}_{[A]}^{(M)} = \{\boldsymbol{b}_{[A]}^k, \ 1 \leq k \leq M\}$ is the M-dimension subset of $\mathbb{H}^M$ that is solution of the Fredholm problem associated with $[A]$:

$$\int_{\Omega} [A(s,s')] \boldsymbol{b}_{[A]}^k(s')ds' = \lambda_{[A]}^k \boldsymbol{b}_{[A]}^k(s), \qquad (19)$$

$$\lambda_{[A]}^1 \geq \lambda_{[A]}^2 \geq \ldots \to 0, \quad \left(\boldsymbol{b}_{[A]}^k, \boldsymbol{b}_{[A]}^\ell\right) = \delta_{k\ell}. \qquad (20)$$

Hence, for all $M \geq 1$, the M-dimensional KL family, $\mathcal{K}^{(M)}$, which is optimal in the sense that it minimizes error ε^2 and that can be searched as the solution of the Fredholm problem defined by Eq. (7), can equivalently be searched as the solution of the following optimization problem:

$$\mathcal{K}^{(M)} = \arg \min_{\mathcal{B}_{[A]}^{(M)}, \, [A] \in \mathcal{S}(\mathbb{R})} \left\{ \varepsilon^2(\mathcal{B}_{[A]}^{(M)}) \right\}. \qquad (21)$$

Hence, it has been shown in Perrin et al. 2012b that when the available information about $\boldsymbol{X}$ is limited to even a relatively small set of independent realizations, as it is the case in this work, it is possible to identify very relevant approximations of $\mathcal{K}^{(M)}$ by searching it as the solution of the following optimization problem:

$$\mathcal{K}^{(M)} \approx \mathcal{B}_{[A(\alpha^*)]}^{(M)}, \qquad (22)$$

$$\begin{cases} [A(\alpha^*)] = [\alpha^*][\widehat{R}] + ([I_4] - [\alpha^*])[\widetilde{R}] \\[2mm] [\alpha^*] = \arg \min_{[\alpha] \in \mathcal{A}} \left\{ \varepsilon^2(\mathcal{B}_{[A(\alpha)]}^{(M)}) \right\}. \end{cases} \qquad (23)$$

$$\mathcal{A} = \{[\alpha], \ | \ [\alpha]_{pq} = \alpha_p \delta_{pq}, 1 \leq p, q \leq 4, \\ (\alpha_1, \ldots, \alpha_4) \in [0,1]^4 \}, \qquad (24)$$

where $[I_4]$ is the (4×4) identity matrix and for all (s, s') in $\Omega \times \Omega$, $1 \leq p, q \leq 4$ and $u = s - s'$:

$$[\widehat{R}(s,s')]_{pq} = \widehat{R}_{pq}(s,s'), \qquad (25)$$

$$[\widetilde{R}(s,s')]_{pq} = \widetilde{R}_{pq}(s,s'), \qquad (26)$$

$$\widehat{R}_{pq}(s,s') = \frac{1}{\nu} \sum_{n=1}^{\nu} x_p^n(s) x_q^n(s'), \qquad (27)$$

$$\widetilde{R}_{pq}(s,s') = \begin{cases} \frac{1}{S+u} \int_0^{S+u} \widehat{R}_{pq}(x, x-u)dx \text{ if } u \in]-S,0], \\ \frac{1}{S-u} \int_0^{S-u} \widehat{R}_{pq}(x+u, x)dx \text{ if } u \in]0, S[, \\ \widehat{R}_{pq}(s,s') \text{ otherwise.} \end{cases} \qquad (28)$$

2.4 Identification of the final projection basis

In Section 2.2, a scaled KL expansion has been presented in order to define projection basis for vector-valued random fields, which allows the minimization of the maximal normalized amplitude of the truncation residue associated with each component of $\boldsymbol{X}$, such that no component is *a priori* defavorized. This scaled KL expansion is however based on the knowledge of covariance function $[R_{\boldsymbol{XX}}]$, which is not available as the maximal information about $\boldsymbol{X}$ is a finite set of independent realizations. In parallel, Section 2.4 has proposed a method to optimize the approximation of the KL basis of $\boldsymbol{X}$ when $\boldsymbol{X}$ is only known from a set of realizations. Therefore, by coupling the two former adaptations of the classical KL expansion, it is possible to identify projection families that are very relevant to condense the statistical information of $\boldsymbol{X}$, while not defavorizing one of its component and even if its covariance function is unknown. For $M \geq 1$, let $\mathcal{U}_\nu^{(M)} = \left\{\boldsymbol{u}^k, 1 \leq k \leq M\right\} \approx \mathcal{F}_{opt}^{(M)}$ be the solution of the optimization problem that combines the two optimization problems defined by Eqs. (14) and (22), that has been identified from the set of ν independent realizations of $\boldsymbol{X}$, $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^\nu\}$.

The projection of $\boldsymbol{X}$ on $\mathcal{U}_\nu^{(M)}$, $\widehat{\boldsymbol{X}}^{(M)}(\mathcal{U}_\nu^{(M)})$ can thus be written:

$$\boldsymbol{X} \approx \widehat{\boldsymbol{X}}^{(M)}(\mathcal{U}_\nu^{(M)}) = \sum_{k=1}^M \boldsymbol{u}^k \eta_k, \ \eta_k = \left(\boldsymbol{X}, \boldsymbol{u}^k\right). \quad (29)$$

Once deterministic projection family $\mathcal{U}_\nu^{(M)}$ has been identified, characterizing the distribution of $\widehat{\boldsymbol{X}}^{(M)}(\mathcal{U}_\nu^{(M)})$ amounts finally to identify the multidimensional distribution of the projection random vector, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_M)$, for which components are *a priori* dependent.

Moreover, the value of truncation parameter M can be identified with respect to an error threshold for error

$$\max_{1 \leq i \leq 4} \frac{\left\| X_i - \widehat{X}_i^{(M)}(\mathcal{U}_\nu^{(M)}) \right\|_2^2}{\|X_i\|_2^2}. \qquad (30)$$

3 INVERSE POLYNOMIAL CHAOS IDENTIFICATION IN HIGH DIMENSION

As presented in Introduction, the second step of the stochastic modeling of the track irregularity random field is the identification of the multidimensional distribution of the KL projection random vector, $\boldsymbol{\eta}$, which is defined by Eq. (29). The available information about this M-dimension random vector is also a set of ν independent realizations, $\{\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^\nu\}$, which is deduced from the realizations of $\boldsymbol{X}$ as:

$$\eta_k^n = \left(\boldsymbol{x}^n, \boldsymbol{u}^k \right), \ \ 1 \le k \le M, \ 1 \le n \le \nu. \tag{31}$$

In this prospect, direct and indirect methods have been developed to identify in inverse the PDF $p_{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$. On the first hand, the direct methods, such as the ones based on the Information Theory and the Maximum Entropy Principle (MEP) (see Jaynes 1963 and Soize 2008), have been introduced to compute $p_{\boldsymbol{\eta}}$ from the only available information about $\boldsymbol{\eta}$. On the other hand, indirect methods aim at constructing PDF $p_{\boldsymbol{\eta}}$ from a transformation $\boldsymbol{t}$ of a chosen random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)$:

$$\boldsymbol{\eta} = \boldsymbol{t}(\boldsymbol{\xi}), \quad p_{\boldsymbol{\eta}} = \mathbb{T}(p_{\boldsymbol{\xi}}). \tag{32}$$

The transformation $\mathbb{T}$ between $p_{\boldsymbol{\eta}}$ and the known and chosen PDF $p_{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ is thus the key step of such indirect methods. One of the most promising indirect methods is currently the polynomial chaos expansion (PCE) method (see Ghanem and Spanos 1990, Ghanem and Spanos 2003, Arnst et al. 2010, Das et al. 2009, Desceliers et al. 2006, Desceliers et al. 2007, Ghanem and Doostan 2006, Marzouk et al. 2007, Perrin et al. 2012a, Soize 2010b, Soize 2010a for further details about the PCE identification in inverse). This technique is based on a direct projection of $\boldsymbol{\eta}$ on a polynomial hilbertian basis $\mathcal{B}_{\text{orth}} = \{\ \psi_j(\boldsymbol{\xi}), \ 1 \le j \ \}$ of all the second-order random vectors with values in $\mathbb{R}^M$, such that:

$$\boldsymbol{\eta} = \sum_{j=1}^{+\infty} \boldsymbol{y}^{(j)} \psi_j(\boldsymbol{\xi}). \tag{33}$$

For practical purposes, this sum is truncated with respect to two truncation parameters, N and N_g, such that:

$$\boldsymbol{\eta} \approx \boldsymbol{\eta}^{\text{chaos}}(N) = \sum_{j=1}^{N} \boldsymbol{y}^{(j)} \psi_j(\xi_1, \ldots, \xi_{N_g}), \tag{34}$$

where projection basis $\left\{ \psi_1(\xi_1, \ldots, \xi_{N_g}), \cdots, \psi_N(\xi_1, \ldots, \xi_{N_g}) \right\}$ is now chosen as the set gathering the N polynomial functions of total degree inferior to p, which are normalized with respect to the PDF $p_{\xi_1, \ldots, \xi_{N_g}}$ of $(\xi_1, \ldots, \xi_{N_g})$.

To identify the PDF of $\boldsymbol{\eta}^{\text{chaos}}(N)$, for given values of N and N_g, we have now to identify the values of the PCE projection coefficients, $\left\{ \boldsymbol{y}^{(j)}, \ 1 \le j \le N \right\}$, from the available independent realizations of $\boldsymbol{\eta}$.

From (Perrin et al. 2012a, Soize 2010b), a good approach to identify such coefficients is to search them as the arguments that maximize the likelihood

of random vector $\boldsymbol{\eta}^{\text{chaos}}(N)$ at the experimental points $\{\boldsymbol{\eta}^1, \cdots, \boldsymbol{\eta}^{\nu^{\text{exp}}}\}$. Advanced algorithms are also presented in (Perrin et al. 2012a, Soize 2010b) to solve such an optimization problem.

At last, to completely identify the PDF of $\boldsymbol{\eta}^{\text{chaos}}(N)$ the choice of the values of N and N_g has to be justified according to a convergence analysis. To this end, a log-error function, $err(N, N_g)$, is introduced to quantify the amplitude of the residue of the PCE truncation, $\boldsymbol{\eta} - \boldsymbol{\eta}^{\text{chaos}}(N)$, such that:

$$err(N, N_g) = \sum_{k=1}^{N_{\eta}} err_k(N, N_g), \tag{35}$$

$$err_k(N, N_g) =$$

$$\int_{\text{BI}_k} \left| log_{10} \left(p_{\eta_k}(x_k) \right) - log_{10} \left(p_{\eta_k^{\text{chaos}}}(x_k) \right) \right| dx_k, \tag{36}$$

where BI_k is the domain bounding the experimental values of η_k, and p_{η_k} and $p_{\eta_k^{\text{chaos}}}$ are the PDFs of η_k and $\eta_k^{\text{chaos}}(N)$ respectively. Truncation parameters N and N_g can therefore be identified with respect to a given error threshold for $err(N, N_g)$.

Finally, once truncation parameters M, N, N_g are identified according to convergence analysis, once PCE projection coefficients $\left\{ \boldsymbol{y}^{(j)}, \ 1 \le j \le N \right\}$ are computed, a complete characterization of the track irregularity random field is given by:

$$\boldsymbol{X} \approx \sum_{k=1}^{M} \boldsymbol{u}^k \sum_{j=1}^{N} y_k^{(j)} \psi_j(\xi_1, \ldots, \xi_{N_g}). \tag{37}$$

For each realization of random vector $(\xi_1, \ldots, \xi_{N_g})$, for which PDF is known, a set of representative and realistic track geometries of length S can finally be generated, which can then be used in any railzay software to characterize the non linear response of the train.

4 VALIDATION OF THE METHOD AND APPLICATION TO A TRACK DATA SET

In Sections 2 and 3, a complete parametrization of the track geometry random field has been presented. This section aims now at showing in what extent such an approach, even if it is based on a relatively small set of independent realizations, can give very relevant results when trying to generate realistic and representative track geometries, which are similar to measured ones from a spectral and statistical point of view.

For confidentiality reasons, the value of S is not given in this paper, and the spatial quantities will be normalized by S in the following.
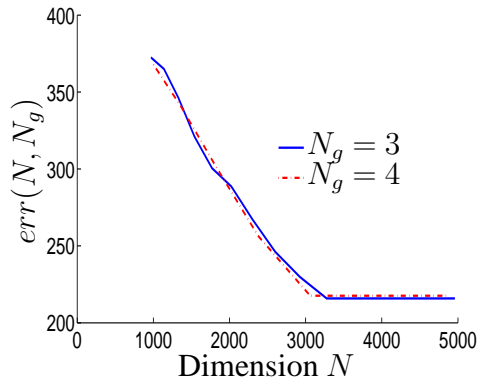
Figure 1: Convergence analysis for the PCE expansion of $\boldsymbol{\eta}$.
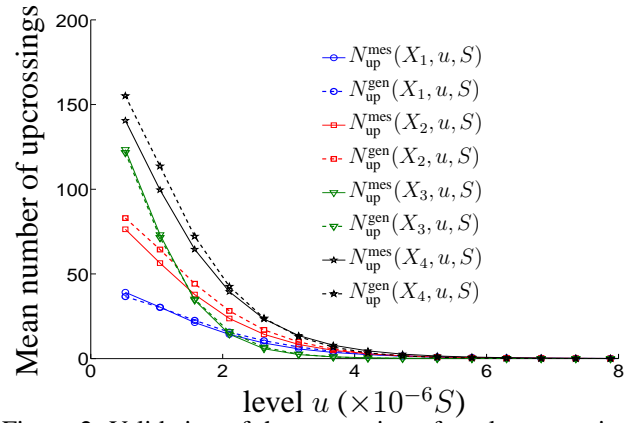

Figure 2: Validation of the generation of track geometries with respect to the number of upcrossings.

Figure 3: Validation of the generation of track geometries with respect to the frequency content.

In this work, $\nu = 414$ track sections of length S are supposed to be known. The scaled KL truncation parameter, M, has to be at least equal to 1900, for the truncation error, defined by Eq. (30) to be lower than 1%. Moreover, $\boldsymbol{\xi}$ is a N_g-dimension random vector, for which components are independent and uniformly distributed between -1 and 1. According to Figure 1, which represents the convergence of error function $err(N, N_g)$ with respect to N and N_g, the values of N and N_g are chosen equal to $3,276$ and 3 respectively. The PCE projection coefficients have then been identified according to the advanced algorithms defined in Perrin et al. 2012a, Soize 2010b.

Hence, $5,00$ track irregularities random fields of total length S are generated from the track stochastic parametrization, which is characterized by Eq. (37). For $1 \leq i \leq 4$, let $N_{\mathrm{up}}^{\mathrm{mes}}(X_i, u, S)$ and $N_{\mathrm{up}}^{\mathrm{gen}}(X_i, u, S)$ be the mean numbers of upcrossings of the level u by X_i over the length S, which have been evaluated from the ν available measured track geometries and from the $5,00$ generated track geometries of length S respectively. In the same manner, for $1 \leq i \leq 4$, let PSD_i^{mes} and $PSD^{\mathrm{gen}}(X_i)$ be the mean power spectral densities of X_i that have been computed from the former ν measured and $5,00$ generated track geometries of length S.

Figures 2 and 3 compare these two quantities for each track irregularity. The results shown in these figures underline therefore the relevance of the proposed approach that is based on a double expansion to characterize the spectral and statistical content of a vector-valued random field, for which the available information is characterized by a finite set of independent realizations.

5 CONCLUSION

A method to identify the statistical characteristics of a non-Gaussian and non-stationary vector-valued random field has been presented in this work, when the available information is a finite set of independent realizations. Based on two original expansions, this methods allows us to generate new realizations of this random field, which are similar to the measured ones,
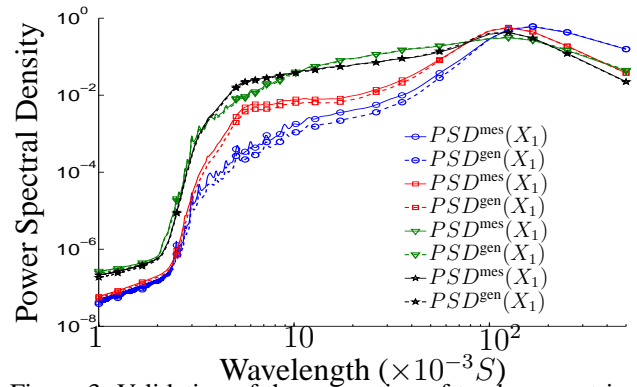
from a frequency and statistical point of view. This method has been applied to a railway track data set, in order to make possible the stochastic analysis of the non-linear response of the trains.

REFERENCES

Arnst, M., R. Ghanem, and C. Soize (2010). Identification of bayesian posteriors for coefficients of chaos expansions. *Journal of Computational Physics 229 (9)*, 3134–3154.

Das, S., R. Ghanem, and S. Finette (2009). Polynomial chaos representation of spatio-temporal random field from experimental measurements. *J. Comput. Phys. 228*, 8726–8751.

Desceliers, C., R. Ghanem, and C. Soize (2006). Maximum likelihood estimation of stochastic chaos representations from experimental data. *Internat. J. Numer. Methods Engrg. 66*, 978–1001.

Desceliers, C., C. Soize, and R. Ghanem (2007). Identification of chaos representations of elastic properties of random media using experimental vibration tests. *Comput. Mech. 39*, 831–838.

Ghanem, R. and A. Doostan (2006). On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data. *J. Comput. Phys. 217*, 63–81.

Ghanem, R. and P. Spanos (1990). Polynomial chaos in stochastic finite elements. *Journal of Applied Mechan-*

ics Transactions of teh ASME 57, 197–202.

Ghanem, R. and P. D. Spanos (2003). *Stochastic Finite Elements: A Spectral Approach, rev. ed.* Dover Publications, New York.

Hansen, P. (1992). Numerical tools for analysis and solution of fredholm integral-equations of the 1st kind. *Inverse problems 8 (6)*, 849–872.

Jaynes, E. T. (1963). Information theory and statistical mechanics. *The Physical Review 106 (4)*, 620–630.

Le Maître, O. and O. Knio (2010). *Spectral Methods for Uncertainty Quantification*. Springer.

Marzouk, Y. M., H. N. Najm, and L. A. Rahn (2007). spectral methods for efficient bayesian solution of inverse problems. *J. Comput. Phys. 224*, 560–586.

Perrin, G., C. Soize, D. Duhamel, and C. Funfschilling (2012a). Identification of polynomial chaos representations in high dimension from a set of realizations. *SIAM J. Sci. Comput. 34(6)*, 2917–2945.

Perrin, G., C. Soize, D. Duhamel, and C. Funfschilling (accepted 2013). Karhunen-loève expansion revisited for vector-valued random fields: scaling, errors and optimal basis. *Journal of Computational Physics*.

Perrin, G., C. Soize, D. Duhamel, and C. Funfschilling (submitted in 2012b). A posteriori error and optimal reduced basis for stochastic processes defined by a finite set of realizations. *SIAM/ASA Journal on Uncertainty Quantification*.

Soize, C. (2008). Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices. *International Journal for Numerical Methods in Engineering 76(10)*, 1583–1611.

Soize, C. (2010a). Generalized probabilistic approach of uncertainties in computational dynamics using random matrices and polynomial chaos decompositions. *Internat. J. Numer. Methods Engrg. 81*, 939–970.

Soize, C. (2010b). Identification of high-dimension polynomial chaos expansions with random coefficients for non-gausian tensor-valued random fields using partial and limited experimental data. *Computer Methods in Applied Mechanics and Engineering 199*, 2150–2164.

Weese, J. (1992). A reliable and fast method for the solution of fredholm integral-equations of the 1st kind based on tikhonov regularization. *Computer physics communications 69*, 99–111.